Bergen, April 3, 1998

Modeling Panels of Intercorrelated Autoregressive Time Series

by

Vidar Hjellvik and Dag Tjøstheim

Department of Mathematics

University of Bergen

5007 Bergen

NORWAY

Summary

We propose a method of modeling panel time series data with both inter- and intra-individual correlation, and of fitting an autoregressive model to such data. Estimates are obtained by a conditional likelihood argument. If there are few observations in each series, the estimates can be dramatically improved by Burg-type estimates taking edge effects into account. The consequences of ignoring the intercorrelation term are analysed. Partial lack of consistency is demonstrated in this situation. Moreover, a break-even point is found for the strength of the intercorrelation, beyond which a conventional estimate, ignoring correlation, will become increasingly inferior. Asymptotic normality of estimators is established, and our results are illustrated on a real data example, where it is seen that choosing the right type of estimate is of crucial importance.

Some key words: Autoregressive; Burg-type estimates; Intercorrelated; Panel data; Time series.

1 Introduction

A quite general linear dynamic model for a panel of time series observations $\{X_{(i)t},\ i = 1,\ldots,n;\ t = 1,\ldots,T\}$ is given by

$$X_{(i)t} = \sum_{j=1}^{p} a_j X_{(i)t-j} + \eta_t + \lambda_i + \beta^\tau W_{(i)t} + \epsilon_{(i)t} \tag{1.1}$$

(see e.g. Hsiao 1986, p. 71). Here t denotes time and i the individual series of the panel. Moreover, $\{W_{(i)t}\}$ is a (possibly) vector series of explanatory variables, η_t represents effects over time influencing all of the series, and similarly λ_i stands for individual effects not taken care of by the explanatory variables. Finally, $\{\epsilon_{(i)t}\}$ are the error terms assumed to be independent identically distributed (iid) in all of the following.

For some reason there seems to be a tradition for removing η_t, thus ignoring the common effects over time and hence effectively the contribution of this term to the intercorrelation across the panel. For example, Diggle, Liang and Zeger (1994) do not include this term. In Baltagi (1995), which contains a recent survey of panel data techniques, inter-individual correlation is only considered briefly for regression models, not for dynamic models. The same is the case in Mátyás and Sevestre (1992). Hsiao (1986), whose chapter 4 is a primary source on dynamic models, in the last sentence of the introduction to this topic writes that "For ease of exposition, we assume that the time specific effects, η_t, do not appear". Similarly, a variable corresponding to η_t is pointedly ignored in the basic paper by Holz-Eatkin et al (1998, p. 1376).

We maintain that the neglect of intercorrelation implied by omitting η_t can in many instances not be justified and may have severe consequences for panels originating, in say, econometrics and biology. We also believe that including η_t is not a trivial extension at the cost of somewhat more burdensome notation. Some of the difficulties are indicated in the scarce literature on the subject. We refer to Diggle and al Wasel (1997, p. 39), who briefly mention this point in a recent discussion on spectral analysis of a panel of time series, and to the earlier papers by Brillinger (1973, 1980) and Bloomfield et al (1983), where common time effects are included.

To focus more sharply at the intercorrelation effect, in this paper we look at the simplified model

$$X_{(i)t} = \sum_{j=1}^{p} a_j X_{(i)t-j} + \eta_t + \epsilon_{(i)t}. \tag{1.2}$$

It is important to understand this relatively simple situation before embarking on models such as (1.1). The variables $\{\eta_t\}$ are generally not assumed to be iid and can be thought of as containing a common mean and also the influence of possible explanatory variables. Of course (1.2) is the panel analogue of a univariate autoregressive time series, whereas (1.1) represents the time series – regression situation.

An illustrative example which can be modeled by (1.2), is depicted in Figure 1, which shows the logarithms of the yearly catches of grey-sided voles over a period of 31 years at 41 different locations of the island of Hokkaido. Clearly the series are intercorrelated, and the geographical area has been chosen so as to minimize individual variations (measured by λ_i in (1.1)) from one catch site to another.

Our primary concern in the present article lies in estimating the part of the dynamic mechanism represented by the autoregressive parameters $a_1, \ldots, a_p$. We are interested in finding good estimates both as $n \to \infty$ with T fixed, and as $T \to \infty$ with n fixed. A related problem with T small ($T \geq 3$) has been treated briefly by Cox and Solomon (1988).

We start with the first order case, where a simple and robust estimate based on a conditional likelihood argument is introduced in Section 2. The main thrust of the paper is to be found in Sections 3 – 6: In section 3 it is shown that the conditional maximum likelihood estimate can be dramatically improved by a Burg-type estimate if T is small and n is large. To our knowledge this type of estimate has not been used before in a panel situation. The consequences of ignoring the intercorrelation term η_t is analysed in Section 4, where a threshold is established for the intercorrelation, beyond which a conventional estimate, ignoring η_t, will be increasingly inferior. Asymptotic normality and an extension to the p-th order autoregressive case are given in Sections

5 and 6. Finally, our results are briefly illustrated on the biological catch data in Section 7.

2 A conditional maximum likelihood estimate

For clarity we first restrict ourselves to $p = 1$ in (1.2). Extensions to an arbitrary p can be found in Section 6. For $p = 1$ the model (1.2) is

$$X_{(i)t} = aX_{(i)t-1} + \eta_t + \epsilon_{(i)t}. \tag{2.1}$$

We assume that observations $\{X_{(i)t}\}$ are available for $i = 1, \ldots, n$ and $t = 1, \ldots, T$, that the $\{\epsilon_{(i)t}\}$ are iid with a density function f_ϵ, and that $|a| < 1$ to guarantee stability. At the moment we make no assumptions about the sequence $\{\eta_t\}$ other than it being independent of $\{\epsilon_{(i)t}\}$. A deterministic sequence $\{\eta_t\}$ would also be allowed.

With the lack of assumptions on $\{\eta_t\}$ unconditional likelihood methods cannot be employed. Even if $\{\eta_t\}$ were to consist of iid random variables, ordinary maximum likelihood arguments cannot in general be used, since if T is small, the intercorrelation introduced by $\{\eta_t\}$ would not be consistently estimated. Conditional on η_t, however, $X_{(i)t}$ and $X_{(j)t}$ are independent for $i \neq j,\ i, j = 1, \ldots, n$. Moreover, denoting by $\mathcal{F}_t^\eta$ the σ-algebra generated by $\{\eta_s,\ s \leq t\}$, by X_t the vector given by $[X_{(1)t}, \ldots, X_{(n)t}]$, and by using a standard Markov argument, the likelihood conditional on $\mathcal{F}_T^\eta$ and the starting value X_1 is given by

$$L(X_2, \ldots, X_T | X_1, \mathcal{F}_T^\eta) = \prod_{t=2}^{T} \prod_{i=1}^{n} f_\epsilon(X_{(i)t} - aX_{(i)t-1} - \eta_t). \tag{2.2}$$

It should be noted that the conditional density of $f(X_1|\mathcal{F}_1^\eta)$ is difficult to evaluate, since, under our general assumptions, $\{X_{(i)t}\}$ is not a stationary process for a fixed i.

If f_ϵ is assumed to be Gaussian with zero mean and variance σ_ϵ^2, then

$$L(X_2, \ldots, X_T | X_1, \mathcal{F}_T^\eta) = (2\pi\sigma_\epsilon^2)^{-n(T-1)/2} \exp\left\{ -\frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^{n} \sum_{t=2}^{T} (X_{(i)t} - aX_{(i)t-1} - \eta_t)^2 \right\}.$$

We can now obtain a consistent estimate of a by letting either n or T, but not necessarily both, tend to infinity. In fact, considering $\{\eta_1, \ldots, \eta_T\}$ to be nuisance parameters, and maximising L with respect to a yields the estimator

$$\tilde{a} = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T-1}(X_{(i)t+1} - \overline{X}_{\cdot t+1})(X_{(i)t} - \overline{X}_{\cdot t})}{\sum_{i=1}^{n}\sum_{t=1}^{T-1}(X_{(i)t} - \overline{X}_{\cdot t})^2} \tag{2.3}$$

where $\overline{X}_{\cdot t} = n^{-1}\sum_{i=1}^{n} X_{(i)t},\ t = 1, \ldots, T$, for $T \geq 2$ and $n \geq 2$. Alternatively, if one wants to avoid the Gaussian assumption, this can be looked at as a conditional least squares estimate. Much of the analysis in the sequel goes through if $\{\epsilon_{(i)t}\}$ is an array of iid random variables satisfying some weak moment conditions.

By (2.1),

$$X_{(i)t} - \overline{X}_{\cdot t} = a(X_{(i)t-1} - \overline{X}_{\cdot t-1}) + \epsilon_{(i)t} - \overline{\epsilon}_{\cdot t}$$

with $\overline{\epsilon}_{\cdot t} = n^{-1}\sum_i \epsilon_{(i)t}$, and it follows that

$$\tilde{a} - a = \frac{\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=1}^{T-1} Y_{(i)t}u_{(i)t+1}}{\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=1}^{T-1} Y_{(i)t}^2} \tag{2.4}$$

where we have used the notation

$$Y_{(i)t} = Y_{(i)t,n} = X_{(i)t} - \overline{X}_{\cdot t} \qquad \text{and} \qquad u_{(i)t} = \epsilon_{(i)t} - \overline{\epsilon}_{\cdot t}. \tag{2.5}$$

It is noted that $\tilde{a}$ is robust in that it does not depend on η_t. Moreover, it is seen from (2.5) and the independence of the $\{\epsilon_{(i)t},\ i = 1, \ldots, n;\ t = 1, \ldots, T\}$ that $\mathrm{E}\,(u_{(i)t}) = 0$ and

$$\mathrm{E}\,(u_{(i)t}u_{(j)s}) = \begin{cases} \sigma_\epsilon^2\left(1 - \frac{1}{n}\right), & i = j, t = s \\ -\frac{1}{n}\sigma_\epsilon^2, & i \neq j, t = s \\ 0, & \text{otherwise.} \end{cases} \tag{2.6}$$

Since $|a| < 1$, $\{Y_{(i)t}\}$ is a stationary process for i (and n) fixed and

$$Y_{(i)t} = \sum_{k=0}^{\infty} a^k u_{(i)t-k}. \tag{2.7}$$

Using the properties of $\{u_{(i)t}\}$, we have that $\mathrm{E}\,(Y_{(i)t}) = 0$ and

$$\mathrm{E}\,(Y_{(i)t}Y_{(j)t}) = \begin{cases} \sigma_Y^2 = \dfrac{\sigma_\epsilon^2\left(1-\frac{1}{n}\right)}{1-a^2}, & i=j \\ -\dfrac{1}{n}\dfrac{\sigma_\epsilon^2}{1-a^2}, & i\neq j \end{cases}, \tag{2.8}$$

whereas for $t > s$ we have $\mathrm{E}\,(Y_{(i)t}Y_{(j)s}) = a^{t-s}\mathrm{E}\,(Y_{(i)t}Y_{(j)t})$. This means that $\{Y_t\} = \{[Y_{(1)t}, \ldots, Y_{(n)t}]\}$ constitutes a first order vector autoregressive process when n is fixed.

Assuming $\mathrm{E}\,(\epsilon_{(i)t}^4) < \infty$ and using the above formulae, some trivial but tedious calculations yield

$$\mathrm{E}\left(\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=1}^{T-1} Y_{(i)t}^2 - \sigma_Y^2\right)^2 \to 0 \tag{2.9}$$

as $nT \to \infty$. Thus, using Chebyshev's inequality for the denominator of (2.4), $n^{-1}(T-1)^{-1}\sum_{i=1}^{n}\sum_{t=1}^{T-1} Y_{(i)t}^2 - \sigma_Y^2 \to 0$ in probability as $nT \to \infty$. Similar but simpler calculations yield

$$\mathrm{E}\left(\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=1}^{T-1} Y_{(i)t}u_{(i)t+1}\right)^2 = \frac{\sigma_Y^2\sigma_u^2}{n(T-1)}\left(1-\frac{1}{n}\right)^{-1} = \frac{\sigma^2\sigma_\epsilon^2}{n(T-1)}\left(1-\frac{1}{n}\right), \tag{2.10}$$

where $\sigma^2 = \sigma_\epsilon^2/(1-a^2)$ is independent of n. Consistency of $\tilde{a}$ follows from (2.4). Moreover, from (2.4), (2.9) and (2.10) it is seen that as $nT \to \infty$,

$$\mathrm{var}\,(\tilde{a}) \sim \frac{1-a^2}{(n-1)(T-1)}. \tag{2.11}$$

If we let $n \to \infty$, by the same argument the estimator

$$\tilde{a}_t = \frac{\sum_{i=1}^{n}(X_{(i)t} - \overline{X}_{\cdot t})(X_{(i)t-1} - \overline{X}_{\cdot t-1})}{\sum_{i=1}^{n}(X_{(i)t-1} - \overline{X}_{\cdot t-1})^2}$$

would be a consistent estimator of a_t in an autoregressive system

$$X_{(i)t} = a_t X_{(i)t-1} + \eta_t + \epsilon_{(i)t}$$

with a time-dependent autoregressive coefficient a_t. But for the model (2.1), $\tilde{\tilde{a}} \doteq (T-1)^{-1}\sum_{t=2}^{T}\tilde{a}_t$ would in general be inferior to $\tilde{a}$, since the latter would work well also for n small (but nT big). Finally, if $T \to \infty$, with some slight abuse of notation, the estimator

$$\tilde{a}_i = \frac{\sum_{t=1}^{T-1} (X_{(i)t+1} - \overline{X}_{(i)\cdot+1})(X_{(i)t} - \overline{X}_{(i)\cdot})}{\sum_{t=1}^{T-1} (X_{(i)t} - \overline{X}_{(i)\cdot})^2}$$

would be a consistent estimator of a_i in an autoregressive system

$$X_{(i)t} = a_i X_{(i)t-1} + \eta_t + \epsilon_{(i)t}$$

with autoregressive coefficients varying from one series to another and with $\overline{X}_{(i)\cdot+1} = (T-1)^{-1} \sum_{t=1}^{T-1} X_{(i)t+1}$ and $\overline{X}_{(i)\cdot} = (T-1)^{-1} \sum_{t=1}^{T-1} X_{(i)t}$.

3 An improved estimator for small T

For the data example mentioned in the introduction, T and n are of about the same size. However, quite often in a panel situation there are many short series. Cox and Solomon, for example, discuss panels of time series each consisting of 3 observations. In such situations it is possible to find a radical improvement of the conditional maximum likelihood estimator $\tilde{a}$.

For ease of computation, in this section the array of variables $\{\epsilon_{(i)t}\}$ will be assumed to be Gaussian, but it will become clear from the derivations that this assumption can be replaced by some appropriate moment conditions. All of the simulations in this paper have been carried out with a Gaussian random number generator.

The reason that $\tilde{a}$ can be improved for T small and n large, is that it is slightly unbalanced at the ends of the data sample. Looking at the expression (2.3) for $\tilde{a}$, it is seen that in the sum in the numerator $X_{(i)1}$ and $X_{(i)T}$ appear once, whilst $X_{(i)s}$, $2 \leq s \leq T-1$ appear twice. On the other hand in the denominator $X_{(i)s}$, $1 \leq s \leq T-1$ appear twice, whereas $X_{(i)T}$ does not appear. This effect is present also for the corresponding conditional likelihood (given X_1) estimator in the single time series case, and it is sometimes corrected for by using a so-called Burg-type estimator (cf. Robinson and

Treitel 1980, appendix 16-2), where $\sum_{t=1}^{T-1}(X_t - \overline{X})^2$ is replaced by $\frac{1}{2}(X_1 - \overline{X})^2 + \frac{1}{2}(X_T - \overline{X})^2 + \sum_{t=2}^{T-1}(X_t - \overline{X})^2$. The panel analogue would be given by

$$\tilde{a}_B = \frac{\dfrac{1}{n(T-1)}\displaystyle\sum_{i=1}^{n}\sum_{t=1}^{T-1}(X_{(i)t+1} - \overline{X}_{\cdot t+1})(X_{(i)t} - \overline{X}_{\cdot t})}{\dfrac{1}{n(T-1)}\displaystyle\sum_{i=1}^{n}\left\{\sum_{t=1}^{T}(X_{(i)t} - \overline{X}_{\cdot t})^2 - \frac{1}{2}(X_{(i)1} - \overline{X}_{\cdot 1})^2 - \frac{1}{2}(X_{(i)T} - \overline{X}_{\cdot T})^2\right\}}. \quad (3.1)$$

Re-introducing the $Y_{(i)t}$- and $u_{(i)t}$-notation, corresponding to (2.4) we have

$$\tilde{a}_B - a = \frac{\dfrac{a}{2n(T-1)}\displaystyle\sum_{i=1}^{n}\left(Y_{(i)1}^2 - Y_{(i)T}^2\right) + \dfrac{1}{n(T-1)}\displaystyle\sum_{i=1}^{n}\sum_{t=1}^{T-1} Y_{(i)t}u_{(i)t+1}}{\dfrac{1}{n(T-1)}\displaystyle\sum_{i=1}^{n}\left\{\sum_{t=1}^{T} Y_{(i)t}^2 - \frac{1}{2}Y_{(i)1}^2 - \frac{1}{2}Y_{(i)T}^2\right\}} \doteq \frac{\hat{A}_1 + \hat{A}_2}{\hat{B}}. \quad (3.2)$$

In univariate time series analysis only the term corresponding to $\hat{A}_2$ contributes to the asymptotic analysis as $T \to \infty$. In the panel case, as $n \to \infty$, both $\hat{A}_1$ and $\hat{A}_2$ contribute. First, as for $\tilde{a}$ it is not difficult to prove that as $n \to \infty$, $\hat{B} \to \sigma^2 = \sigma_\epsilon^2/(1-a^2)$ and $\hat{A}_i \to 0, i = 1, 2$ in probability. Using a Taylor expansion argument it follows that

$$\mathrm{var}\,(\tilde{a}_B) \sim \frac{\mathrm{E}\,(\hat{A}_1^2) + \mathrm{E}\,(\hat{A}_2^2) + 2\mathrm{E}\,(\hat{A}_1\hat{A}_2)}{\sigma^4}. \quad (3.3)$$

Here, using (2.10), $\mathrm{E}\,(\hat{A}_2^2) \sim n^{-1}(T-1)^{-1}\sigma^2\sigma_\epsilon^2$ as $n \to \infty$, and it remains to evaluate $\mathrm{E}\,(\hat{A}_1^2)$ and $\mathrm{E}\,(\hat{A}_1\hat{A}_2)$. In the derivation we utilize the relationship

$$Y_{(i)t} = a^{t-s}Y_{(i)s} + \sum_{k=s+1}^{t} a^{t-k}u_{(i)k},\ t > s. \quad (3.4)$$

Note that

$$\mathrm{E}\,(\hat{A}_1^2) = \frac{a^2}{4n^2(T-1)^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\mathrm{E}\,(Y_{(i)1}^2Y_{(j)1}^2) - 2\mathrm{E}\,(Y_{(i)1}^2Y_{(j)T}^2) + \mathrm{E}\,(Y_{(i)T}^2Y_{(j)T}^2)\right\}. \quad (3.5)$$

Using the Gaussian assumption on $\{\epsilon_{(i)t}\}$,

$$\mathrm{E}\,(Y_{(i)1}^4) = \mathrm{E}\,(Y_{(i)T}^4) = 3\sigma_Y^4. \quad (3.6)$$

Moreover, for $i \neq j$, using (3.4) with $t = 2$ and $s = 1$,

$$\begin{aligned}
&\mathrm{E}\,\left(Y_{(i)2}^2Y_{(j)2}^2\right) \\
&= \mathrm{E}\,\left\{\left(a^2Y_{(i)1}^2 + 2aY_{(i)1}u_{(i)2} + u_{(i)2}^2\right)\left(a^2Y_{(j)1}^2 + 2aY_{(j)1}u_{(j)2} + u_{(j)2}^2\right)\right\} \\
&= a^4\mathrm{E}\,\left(Y_{(i)1}^2Y_{(j)1}^2\right) + 2a^2\sigma_Y^2\sigma_u^2 + \sigma_u^4 + \frac{4a^2}{(n-1)^2}\sigma_Y^2\sigma_u^2. \quad (3.7)
\end{aligned}$$

Using (2.7), it follows that $\mathrm{E}\,(Y_{(i)2}^2 Y_{(j)2}^2) = E(Y_{(i)1}^2 Y_{(j)1}^2) = \mathrm{E}\,(Y_{(i)T}^2 Y_{(j)T}^2)$, and by (2.6), (2.8) and (3.7)

$$\mathrm{E}\,(Y_{(i)2}^2 Y_{(j)2}^2) = \frac{2a^2\sigma_Y^2\sigma_u^2 + \sigma_u^4}{1-a^4} + O(n^{-2}) = \sigma_Y^4 + O(n^{-2}). \tag{3.8}$$

Similarly, with $t = T$ and $s = 1$ in (3.4), and using (3.6), (3.8) and the independence of $\{u_{(j)k}, k \geq 2\}$ from $Y_{(j)1}$ we obtain

$$\begin{aligned}
\mathrm{E}\,\left(Y_{(i)1}^2 Y_{(j)T}^2\right) &= \mathrm{E}\,\left\{Y_{(i)1}^2\left(a^{T-1}Y_{(j)1} + \sum_{k=2}^{T} a^{T-k}u_{(j)k}\right)^2\right\} \\
&= \mathrm{E}\,\left\{Y_{(i)1}^2\left(a^{2(T-1)}Y_{(j)1}^2 + \sum_{k=2}^{T} a^{2(T-k)}u_{(j)k}^2\right)\right\} \\
&= \begin{cases} \left\{3a^{2(T-1)} + (1-a^2)\sum\limits_{k=2}^{T} a^{2(T-k)}\right\}\sigma_Y^4, & i = j \\ \left\{a^{2(T-1)} + (1-a^2)\sum\limits_{k=2}^{T} a^{2(T-k)}\right\}\sigma_Y^4 + O(n^{-2}), & i \neq j \end{cases} \\
&= \begin{cases} \left(2a^{2(T-1)} + 1\right)\sigma_Y^4, & i = j \\ \sigma_Y^4 + O(n^{-2}), & i \neq j \end{cases}.
\end{aligned} \tag{3.9}$$

It follows from (3.8) and (3.9) that for $i \neq j$ the terms of (3.5) cancel up to order $O(n^{-2})$, and for $i = j$, by inserting (3.6) and (3.9) in (3.5),

$$\mathrm{E}\,(\hat{A}_1^2) = \frac{a^2\left(1 - a^{2(T-1)}\right)\sigma_Y^4}{n(T-1)^2} + O(n^{-2}) \sim \frac{a^2\left(1 - a^{2(T-1)}\right)\sigma^4}{n(T-1)^2} \tag{3.10}$$

as $n \to \infty$.

Finally, for the covariance term $\mathrm{E}\,(\hat{A}_1\hat{A}_2)$ of (3.3) we have

$$\begin{aligned}
\mathrm{E}\,(\hat{A}_1\hat{A}_2) &= \frac{a}{2n^2(T-1)^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T-1}\mathrm{E}\,\left\{\left(Y_{(i)1}^2 - Y_{(i)T}^2\right)Y_{(j)t}u_{(j)t+1}\right\} \\
&= \frac{-a}{2n^2(T-1)^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=1}^{T-1}\mathrm{E}\,\left(Y_{(i)T}^2 Y_{(j)t}u_{(j)t+1}\right).
\end{aligned}$$

Using (2.6), (2.7), (2.8) and (3.4) with $t = T$ and $s = t$,

$$\mathrm{E}\,\left(Y_{(i)T}^2 Y_{(j)t}u_{(j)t+1}\right) = \mathrm{E}\,\left\{\left(a^{T-t}Y_{(i)t} + \sum_{k=t+1}^{T} a^{T-k}u_{(i)k}\right)^2 Y_{(j)t}u_{(j)t+1}\right\}$$

$$= \mathrm{E}\left(2a^{T-t}Y_{(i)t}Y_{(j)t}a^{T-t-1}u_{(i)t+1}u_{(j)t+1}\right)$$

$$= \begin{cases} 2a^{2(T-t)-1}(1-a^2)\sigma_Y^4, & i=j \\ \dfrac{2}{(n-1)^2}a^{2(T-t)-1}(1-a^2)\sigma_Y^4, & i\neq j \end{cases}$$

so that, as $n\to\infty$,

$$\mathrm{E}\,(\hat{A}_1\hat{A}_2) \sim \frac{-a^2\left(1-a^{2(T-1)}\right)\sigma^4}{n(T-1)^2}. \tag{3.11}$$

Inserting (3.10), (3.11) and $\mathrm{E}\,(\hat{A}_2^2)\sim n^{-1}(T-1)^{-1}(1-a^2)\sigma^4$ into (3.3), we obtain

$$\mathrm{var}\,(\tilde{a}_B) \sim \frac{(1-a^2)(T-1)-a^2\left(1-a^{2(T-1)}\right)}{n(T-1)^2} = \frac{T-1-Ta^2+a^{2T}}{n(T-1)^2}. \tag{3.12}$$

Table 1 shows, for various combinations of T, n and a, the simulated value $n(T-1)\widehat{\mathrm{var}}\,(\tilde{a}_B)$ and the ratio $\widehat{\mathrm{var}}\,(\tilde{a}_B)/\mathrm{var}\,(\tilde{a}_B)$, where $\mathrm{var}\,(\tilde{a}_B)$ is given by (3.12). It is seen that (3.12) results in a slight under-estimation for $n=128$, but this effect has all but disappeared for $n=1024$.

By (2.11) and (3.12), as $n\to\infty$,

$$\mathrm{var}\,(\tilde{a}) \sim \frac{1-a^2}{n(T-1)} \sim \mathrm{var}\,(\tilde{a}_B) + \frac{a^2\left(1-a^{2(T-1)}\right)}{n(T-1)^2}$$

and hence

$$\frac{\mathrm{var}\,(\tilde{a}_B)}{\mathrm{var}\,(\tilde{a})} \sim 1 - \frac{a^2\left(1-a^{2(T-1)}\right)}{(T-1)(1-a^2)}. \tag{3.13}$$

If T is allowed to increase, $\mathrm{var}\,(\tilde{a}_B)/\mathrm{var}\,(\tilde{a})\to 1$, whereas for T small and $|a|$ large, $\tilde{a}_B$ is much better than $\tilde{a}$. For $T=2$, for example,

$$\frac{\mathrm{var}\,(\tilde{a}_B)}{\mathrm{var}\,(\tilde{a})} \sim 1-a^2$$

as $n\to\infty$. The asymptotic relationship (3.13) stands in startling contrast to traditional univariate time series analysis, where T typically is so large that the Burg estimator and the conditional likelihood estimator very nearly have the same variance. In Figure 2, $\tilde{a}_B$ and $\tilde{a}$ are compared for $T=2$ and $n=128$. The simulated mean square error is given in Table 2, where we also see that the ratio $\widehat{\mathrm{MSE}}\,(\tilde{a}_B)/\widehat{\mathrm{MSE}}\,(\tilde{a})$

is close to the ratio $\mathrm{var}(\tilde{a}_B)/\mathrm{var}(\tilde{a})$ given by (3.13). The simulated bias squared is typically of order 10^{-5}, so the mean square error is practically equal to the variance here. It is seen that a can be accurately estimated with only 2 observations for each of the series of the panel. We also note from Figure 2 that for $|a| = 0.9$, $|\tilde{a}| > 1$ in some cases. On the other hand, an application of the Schwarz inequality in (3.1) and some simple algebra show that $\tilde{a}_B < 1$ always.

4 Ignoring intercorrelation: consequences and a break-even point

If it is assumed that the time series $\{X_{(i)t}\}$, $i = 1, 2, \ldots$ are independent, so that (2.1) simplifies to

$$X_{(i)t} = aX_{(i)t-1} + b + \epsilon_{(i)t}, \tag{4.1}$$

where b is a constant into which a possible non-zero mean has been absorbed, then the need to condition on $\mathcal{F}_T^{\eta}$ in (2.2) disappears, and the conditional likelihood given X_1 is

$$L(X_2, \ldots, X_T | X_1) = \prod_{t=2}^{T} \prod_{i=1}^{n} f_\epsilon(X_{(i)t} - aX_{(i)t-1} - b).$$

The corresponding conditional maximum likelihood estimator of a is given by

$$\hat{a} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T-1} (X_{(i)t+1} - \overline{X}_{\cdot\cdot,1})(X_{(i)t} - \overline{X}_{\cdot\cdot,0})}{\sum_{i=1}^{n} \sum_{t=1}^{T-1} (X_{(i)t} - \overline{X}_{\cdot\cdot,0})^2} \tag{4.2}$$

with $\overline{X}_{\cdot\cdot,j} = n^{-1}(T-1)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T-1} X_{(i)t+j}$, $j = 0, 1$. Arguing as in Section 2, $\hat{a}$ is consistent under the model (4.1) as $nT \to \infty$ and

$$\mathrm{var}(\hat{a}) \sim \frac{1 - a^2}{n(T-1)}. \tag{4.3}$$

For a small T it can be improved at the edges as in Section 3, resulting in an estimator $\hat{a}_B$ analogous to $\tilde{a}_B$. Note that under (4.1) an alternative way of proceeding would be to replace $\hat{a}$ by the unconditional maximum likelihood estimate obtained by including the marginal distribution of X_1 in the likelihood maximization.

Whereas $\tilde{a}$ and $\tilde{a}_B$ are consistent under (4.1) *and* the more general model (2.1) including intercorrelation, this is not so for $\hat{a}$ and $\hat{a}_B$. Indeed, assuming that (2.1) is true, from (4.2) we obtain

$$\hat{a} = a + \frac{\frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T-1} (X_{(i)t} - \overline{X}_{\cdot\cdot,0})(\epsilon_{(i)t+1} - \overline{\epsilon}_{\cdot\cdot,1}) + \frac{1}{T-1} \sum_{t=1}^{T-1} (\overline{X}_{\cdot t} - \overline{X}_{\cdot\cdot,0})(\eta_{t+1} - \overline{\eta}_{\cdot,1})}{\frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T-1} (X_{(i)t} - \overline{X}_{\cdot\cdot,0})^2} \tag{4.4}$$

where $\overline{\epsilon}_{\cdot\cdot,1} = n^{-1}(T-1)^{-1} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \epsilon_{(i)t+1}$ and $\overline{\eta}_{\cdot,1} = (T-1)^{-1} \sum_{t=1}^{T-1} \eta_{t+1}$. For a fixed T, the term $(T-1)^{-1} \sum_t (\overline{X}_{\cdot t} - \overline{X}_{\cdot\cdot,0})(\eta_{t+1} - \overline{\eta}_{\cdot,1})$ will not converge to zero in probability as $n \to \infty$, and even if T tends to infinity, it is not unproblematic to analyse (4.4) unless further assumptions are made about $\{\eta_t\}$ to obtain stationarity of $\{X_{(i)t}\}$.

One may argue that the situations represented by (2.1) and (4.1) are too simple, and that the inconsistency of $\hat{a}$ under (2.1) is of minor practical importance, as typically more complicated models tend to be used. However, for the entirely general model (1.1), if η_t is omitted — as it usually is — the estimates of $\hat{a}_1, \ldots, \hat{a}_p$ conventionally employed (see e.g. Hsiao 1986, Ch. 4) would generally not be consistent unless $\eta_t \equiv$ constant. For example, the much used differencing technique yields a p-th order autoregressive model for the differences $Z_{(i)t} = X_{(i)t} - X_{(i)t-1}$,

$$Z_{(i)t} = \sum_{j=1}^{p} a_j Z_{(i)t-1} + \eta_t - \eta_{t-1} + \epsilon_{(i)t} - \epsilon_{(i)t-1} + \beta^\tau (W_{(i)t} - W_{(i)t-1}).$$

Estimators of $a_1, \ldots, a_p$ analogous to those defined in Hsiao (1986, Ch. 4) would not be consistent as $n \to \infty$ for T fixed if intercorrelation is present, whereas estimators analogous to $\tilde{a}$ in Section 3, based on conditional likelihood arguments, would.

Coming back to the simpler picture presented by (2.1) and (4.1), we will introduce assumptions on $\{\eta_t\}$ so that $\hat{a}$ is consistent under (2.1) as $T \to \infty$. One reason this is of interest, is that as $T \to \infty$ for n fixed, according to (4.3), $\mathrm{var}\,(\hat{a}) \sim$ $n^{-1}(T-1)^{-1}(1-a^2)$ under model (4.1), whereas from (2.11), as $T \to \infty$ for n fixed, $\mathrm{var}\,(\tilde{a}) \sim (n-1)^{-1}(T-1)^{-1}(1-a^2)$ which holds both under (4.1) and (2.1). Thus if the series $\{X_{(i)t}\}$, $i = 1, 2, \ldots$ are independent, as $T \to \infty$,

$$\frac{\mathrm{var}\,(\hat{a})}{\mathrm{var}\,(\tilde{a})} \sim \frac{n-1}{n}$$

and for small n, $\hat{a}$ would be distinctly preferable to $\tilde{a}$. (Of course $\mathrm{var}\,(\hat{a}_B) \sim \mathrm{var}\,(\hat{a})$ and $\mathrm{var}\,(\tilde{a}_B) \sim \mathrm{var}\,(\tilde{a})$ in this case). A natural question is: can $\hat{a}$ be better than $\tilde{a}$ also when intercorrelation is present, and if so, when?

The needed assumptions on η_t to obtain consistency of $\hat{a}$ as $T \to \infty$ under (2.1) are quite restrictive; in fact the η_t-variables should be iid . It is *not* sufficient that $\{\eta_t\}$ is stationary, because if $\{\eta_t\}$ is autocorrelated, $\hat{a}$ would be a consistent estimator of $\mathrm{corr}\,(X_{(i)t}, X_{(i)t-1})$ but not of a. Depending on one's point of view the restriction to iid variables may be quite serious. A recent development within multiple time series modeling concerns so-called factor models (Forni and Reichlin 1997), where much emphasis is put on a *stationary* analogue of $\{\eta_t\}$ describing common dynamics due to external economic factors.

Assuming in addition that the variables $\{\eta_t\}$ has a finite second moment and iterating in (2.1), we obtain corresponding to (2.7),

$$X_{(i)t} = \sum_{k=0}^{\infty} a^k(\eta_{t-k} + \epsilon_{(i)t-k}). \tag{4.5}$$

Using that $\{\eta_t\}$ and $\{\epsilon_{(i)t}\}$, $i = 1, 2, \ldots$, are independent sequences of iid variables, it follows at once that

$$\mathrm{E}\,(X_{(i)t}) = (1-a)^{-1}\mathrm{E}\,(\eta_t), \tag{4.6}$$

$$\mathrm{var}\,(X_{(i)t}) = (1-a^2)^{-1}(\sigma_\eta^2 + \sigma_\epsilon^2), \quad \mathrm{cov}\,(X_{(i)t}, X_{(j)t}) = (1-a^2)^{-1}\sigma_\eta^2,\ i \neq j, \tag{4.7}$$

and consequently

$$\mathrm{corr}\,(X_{(i)t}, X_{(j)t}) \doteq \rho = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\epsilon^2}. \tag{4.8}$$

Moreover, for $t > s$, $\mathrm{cov}\,(X_{(i)t}, X_{(j)s}) = a^{t-s}\mathrm{cov}\,(X_{(i)t}, X_{(j)t})$.

Using the above formulae and (4.4), it is not difficult to show that $\hat{a}$ is consistent under (2.1) as $T \to \infty$ under the added assumption that 4-th moments exist. Furthermore, we are in a position to address the question of the relative efficiency of $\tilde{a}$ and $\hat{a}$ as $T \to \infty$. Defining

$$\hat{C}_i = \frac{1}{T-1}\sum_{t=1}^{T-1}(X_{(i)t} - \overline{X}_{\cdot\cdot,0})(\eta_{t+1} - \overline{\eta}_{\cdot,1} + \epsilon_{(i)t+1} - \overline{\epsilon}_{\cdot\cdot,1}),$$

we can write (4.4) as

$$\hat{a} - a = \frac{\frac{1}{n}\sum_{i=1}^{n}\hat{C}_i}{\frac{1}{n(T-1)}\sum_{i=1}^{n}\sum_{t=1}^{T-1}(X_{(i)t} - \overline{X}_{\cdot\cdot,0})^2}$$

and using a Taylor expansion argument, as $T \to \infty$,

$$\operatorname{var}(\hat{a}) \sim \frac{\operatorname{var}\left(\frac{1}{n}\sum_{i=1}^{n}\hat{C}_i\right)}{\sigma_X^4} \sim \frac{\mathrm{E}\,(\hat{C}_1^2) + (n-1)\mathrm{E}\,(\hat{C}_1\hat{C}_2)}{n\sigma_X^4}. \tag{4.9}$$

By (4.5)-(4.8),

$$\begin{aligned} \mathrm{E}\,(\hat{C}_1^2) &\sim \frac{1}{(T-1)^2}\sum_{t=1}^{T-1}\mathrm{E}\left[(X_{(1)t} - \overline{X}_{\cdot\cdot,0})^2\left\{(\eta_{t+1} - \overline{\eta}_{\cdot,1})^2 + (\epsilon_{(1)t+1} - \overline{\epsilon}_{\cdot\cdot,1})^2\right\}\right] \\ &\sim \frac{\sigma_X^2(\sigma_\eta^2 + \sigma_\epsilon^2)}{T-1} = \frac{(1-a^2)\sigma_X^4}{T-1} \end{aligned}$$

and

$$\begin{aligned} \mathrm{E}\,(\hat{C}_1\hat{C}_2) &\sim \frac{1}{(T-1)^2}\sum_{t=1}^{T-1}\mathrm{E}\left\{(X_{(1)t} - \overline{X}_{\cdot\cdot,0})(X_{(2)t} - \overline{X}_{\cdot\cdot,0})(\eta_{t+1} - \overline{\eta}_{\cdot,1})^2)\right\} \\ &\sim \frac{\operatorname{cov}(X_{(1)t}, X_{(2)t})\sigma_\eta^2}{T-1} = \frac{\rho\sigma_X^2\sigma_\eta^2}{T-1} = \frac{\rho^2(1-a^2)\sigma_X^4}{T-1}. \end{aligned}$$

Inserting in (4.9), it follows that

$$\operatorname{var}(\hat{a}) \sim \operatorname{var}(\hat{a}_B) \sim \frac{(1-a^2)\{1 + (n-1)\rho^2\}}{n(T-1)} \tag{4.10}$$

and comparing to (2.11), which is valid irrespective of whether intercorrelation is present or not,

$$\frac{\operatorname{var}(\hat{a})}{\operatorname{var}(\tilde{a})} \sim \frac{\operatorname{var}(\hat{a}_B)}{\operatorname{var}(\tilde{a}_B)} \sim \frac{\{1 + (n-1)\rho^2\}(n-1)}{n} \tag{4.11}$$

as $T \to \infty$. It is seen that $\rho = 1/(n-1)$ marks a threshold such that asymptotically as $T \to \infty$,

$$\operatorname{var}(\hat{a}) \le \operatorname{var}(\tilde{a}) \quad \text{if and only if} \quad \rho \le \frac{1}{n-1}. \tag{4.12}$$

Since ρ can be estimated consistently as $T \to \infty$, this gives a simple rule for choosing between $\hat{a}$ and $\tilde{a}$ when T is large. When T is small, $\tilde{a}$, or rather $\tilde{a}_B$, should always be preferred due to the lack of consistency of $\hat{a}$ and $\hat{a}_B$ in this case.

In Figure 3 and Table 3, $\tilde{a}_B$ and $\hat{a}_B$ are compared for $n = 2$, for which (4.12) is always fulfilled, and $n = 4$. For $nT = 20000$ the simulated bias squared is of order 10^{-7}, so the simulated mean square errors are practically equal to the simulated variances, and they are very close to the theoretical values given by (2.11) and (4.10), which for $nT = 20000$ are given in parentheses in Table 3.

At the break-even point $\rho = 1/(n-1)$ the simulated variance of both $\hat{a}_B$ and $\tilde{a}_B$ should be close to $\mathrm{var}\,(\tilde{a}_B)$ given by (2.11). Table 4 shows that this is indeed the case. In the table simulated results are compared to (2.11) for $n = 2, 3$ and 4, $nT = 20000$ and $a = 0.5$. Even though these are asymptotic results, they seem to hold for T quite small, as can be seen from Table 5. However, the bias of $\hat{a}_B$ is typically about three times that of $\tilde{a}_B$. For T large the bias squared is small compared to the variance, but for T small it is substantial for $\hat{a}_B$.

For T large and $n = 2$, $\hat{a}$ and $\hat{a}_B$ should always be used. However, from (4.11) and (4.12) it is clear that as n is increasing, there is almost nothing to gain by using $\hat{a}$ and $\hat{a}_B$, but there is a lot to lose in the intercorrelated case. In Figure 4 and Table 6, $\hat{a}_B$ is compared to $\tilde{a}_B$ for $n = 128$ and $\rho = 0.5$, which gives $\mathrm{var}\,(\hat{a}_B)/\mathrm{var}\,(\tilde{a}_B) \sim 32.5$ in (4.11). From Figure 4 we also see that $\hat{a}_B$ is biased, and for T small the bias is severe. This is not surprising in view of the inconsistency of $\hat{a}$ for T small. Similar results can be expected for the more general model (1.1) for estimates that do not take the intercorrelation effect into account.

5 Asymptotic distribution

If T is allowed to tend to infinity, strong mixing arguments can be used to prove asymptotic normality for both $\tilde{a}$ and $\hat{a}$, with no assumptions on $\{\eta_t\}$ in the case of $\tilde{a}$. If T is fixed, and n tends to infinity, the problem can be reduced to the central limit theorem for iid random variables, as will now be shown.

We shall start with the estimator $\tilde{a}$. Since the denominator of (2.4) converges to σ^2 in probability as $n \to \infty$, it is enough to consider the numerator

$$\hat{A}_2 = \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T-1} u_{(i)t+1} Y_{(i)t}.$$

Inserting $u_{(i)t+1} = \epsilon_{(i)t+1} - \overline{\epsilon}_{\cdot t+1}$ we have

$$\begin{aligned} \hat{A}_2 &= \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \epsilon_{(i)t+1} Y_{(i)t} - \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \overline{\epsilon}_{\cdot t+1} Y_{(i)t} \\ &= \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \epsilon_{(i)t+1} Y_{(i)t} \end{aligned}$$

since $n^{-1} \sum_{i=1}^{n} Y_{(i)t} = \overline{X}_{\cdot t} - \overline{X}_{\cdot t} = 0$. Let $Z_{(i)t}$ be defined by

$$Z_{(i)t} = a Z_{(i)t-1} + \epsilon_{(i)t}$$

so that

$$Z_{(i)t} = \sum_{k=0}^{\infty} a^k \epsilon_{(i)t-k}.$$

Comparing with (2.7), it is seen that $Y_{(i)t} = Z_{(i)t} - \overline{Z}_{\cdot t}$ and hence

$$\begin{aligned} \hat{A}_2 &= \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \epsilon_{(i)t+1} Y_{(i)t} \\ &= \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \epsilon_{(i)t+1} Z_{(i)t} - \frac{1}{T-1} \sum_{t=1}^{T-1} \overline{\epsilon}_{\cdot t+1} \overline{Z}_{\cdot t}. \end{aligned}$$

Here

$$\mathrm{E}\,(\overline{\epsilon}^2_{\cdot t+1} \overline{Z}^2_{\cdot t}) = \frac{1}{n^4} \mathrm{E} \left(\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \sum_{i_3=1}^{n} \sum_{i_4=1}^{n} \epsilon_{(i_1)t+1} \epsilon_{(i_2)t+1} Z_{(i_3)t} Z_{(i_4)t} \right) = O(n^{-2})$$

due to the independence assumption on the array $\{\epsilon_{(i)t}\}$. On the other hand, by standard arguments,

$$\mathrm{E} \left\{ \left(\frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=1}^{T-1} \epsilon_{(i)t+1} Z_{(i)t} \right)^2 \right\} = O(n^{-1}),$$

and it follows that it is sufficient to consider

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T-1} \sum_{t=1}^{T-1} \epsilon_{(i)t+1} Z_{(i)t} \doteq \frac{1}{n} \sum_{i=1}^{n} D_{i,T}$$

where the random variables $\{D_{1,T}, D_{2,T}, \ldots\}$ are independent, and the asymptotic normality of $\tilde{a}$ then follows by an application of the ordinary central limit theorem on $\{D_{i,T},\ i = 1, 2, \ldots\}$.

A similar reduction to iid random variables can be achieved for the estimator $\tilde{a}_B$. By (3.2) and the above, it is sufficient to look at the term

$$\begin{aligned}\hat{A}_1 &= \frac{a}{2n(T-1)}\sum_{i=1}^{n}\left(Y_{(i)1}^2 - Y_{(i)T}^2\right)\\ &= \frac{a}{2n(T-1)}\sum_{i=1}^{n}\left(Z_{(i)1}^2 - Z_{(i)T}^2\right) - \frac{a}{2(T-1)}\left(Z_{\cdot 1}^2 - Z_{\cdot T}^2\right).\end{aligned}$$

Here

$$\mathrm{E}\left\{\left(Z_{\cdot 1}^2 - Z_{\cdot T}^2\right)^2\right\} = \frac{1}{n^4}\mathrm{E}\left\{\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\sum_{i_4=1}^{n}(Z_{(i_1)1}Z_{(i_2)1}Z_{(i_3)1}Z_{(i_4)1}\right.$$

$$\left. - \quad Z_{(i_1)1}Z_{(i_2)1}Z_{(i_3)T}Z_{(i_4)T} - Z_{(i_1)T}Z_{(i_2)T}Z_{(i_3)1}Z_{(i_4)1} + Z_{(i_1)T}Z_{(i_2)T}Z_{(i_3)T}Z_{(i_4)T})\right\} = O(n^{-2})$$

by the independence and zero-mean properties of the array $\{Z_{(i)t}\}$. On the other hand

$$\begin{aligned}&\mathrm{E}\left[\left\{\frac{1}{n}\sum_{i=1}^{n}\left(Z_{(i)1}^2 - Z_{(i)T}^2\right)\right\}^2\right]\\ &= \frac{1}{n^2}\mathrm{E}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}(Z_{(i)1}^2Z_{(j)1}^2 - Z_{(i)1}^2Z_{(j)T}^2 - Z_{(i)T}^2Z_{(j)1}^2 + Z_{(i)T}^2Z_{(j)T}^2)\right\}\\ &= \frac{2}{n}\left\{\mathrm{E}\,(Z_{(i)1}^4) - \mathrm{E}\,(Z_{(i)1}^2Z_{(i)T}^2)\right\} = O(n^{-1}),\end{aligned}$$

and hence the dominating term of the numerator $\hat{A}_1 + \hat{A}_2$ of (3.2) is given by

$$\frac{1}{n}\sum_{i=1}^{n}\left\{\frac{a}{2(T-1)}\left(Z_{(i)1}^2 - Z_{(i)T}^2\right) + \frac{1}{T-1}\sum_{t=1}^{T-1}\epsilon_{(i)t+1}Z_{(i)t}\right\} \doteq \frac{1}{n}\sum_{i=1}^{n}F_{i,T}$$

and the asymptotic normality of $\tilde{a}_B$ follows from an application of the ordinary central limit theorem to the sequence of independent random variables $\{F_{i,T},\ i = 1, 2, \ldots\}$. A simulation experiment, illustrating the distributional results for $T = 2$, is presented in Figure 5.

6 The autoregressive model of order p

We start by writing (1.2) in vector notation:

$$X_{(i)t} = a^{\tau}x_{(i)t-1} + \eta_t + \epsilon_{(i)t} \tag{6.1}$$

where $a^\tau = [a_1, \ldots, a_p]$ and $x_{(i)t}^\tau = [X_{(i)t}, X_{(i)t-1}, \ldots, X_{(i)t-p+1}]$. The conditional likelihood argument of Section 2 can now be repeated to obtain

$$\begin{aligned} \tilde{a}^\tau &= \frac{1}{n(T-p)} \sum_{i=1}^{n} \sum_{t=p}^{T-1} (X_{(i)t+1} - \overline{X}_{\cdot t+1})(x_{(i)t} - \overline{x}_{\cdot t})^\tau \\ &\quad \times \left\{ \frac{1}{n(T-p)} \sum_{i=1}^{n} \sum_{t=p}^{T-1} (x_{(i)t} - \overline{x}_{\cdot t})(x_{(i)t} - \overline{x}_{\cdot t})^\tau \right\}^{-1} . \end{aligned} \tag{6.2}$$

Introducing $Y_{(i)t}$ and $u_{(i)t}$ as before and letting $y_{(i)t}^\tau = [Y_{(i)t}, Y_{(i)t-1}, \ldots, Y_{(i)t-p+1}]$, corresponding to (2.4),

$$\tilde{a}^\tau - a^\tau = \frac{1}{n(T-p)} \sum_{i=1}^{n} \sum_{t=p}^{T-1} u_{(i)t+1} y_{(i)t}^\tau \left\{ \frac{1}{n(T-p)} \sum_{i=1}^{n} \sum_{t=p}^{T-1} y_{(i)t} y_{(i)t}^\tau \right\}^{-1} . \tag{6.3}$$

To obtain an expression for the covariance matrix $\Gamma_y = \mathrm{E}\,(y_{(i)t} y_{(i)t}^\tau)$ of $y_{(i)t}$ note that

$$y_{(i)t} = A y_{(i)t-1} + u_{(i)t} 1_v, \tag{6.4}$$

where A is the $p \times p$ matrix defined by

$$A = \begin{bmatrix} a_1 & a_2 & \ldots & a_{p-1} & a_p \\ 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}, \tag{6.5}$$

and 1_v is a $p-$dimensional vector having 1 as its first entry and zeros otherwise. Assuming that the characteristic polynomial $z^p - \sum_{i=1}^{p} a_i z^{p-i}$ has its zeros inside the unit circle, it follows by (6.4) that

$$y_{(i)t} = \sum_{k=0}^{\infty} A^k 1_v u_{(i)t-k},$$

and hence corresponding to (2.8),

$$\mathrm{E}\,(y_{(i)t} y_{(j)t}^\tau) = \begin{cases} \Gamma_y = \left(1 - \dfrac{1}{n}\right) \Gamma, & i = j \\ -\dfrac{1}{n}\Gamma, & i \neq j \end{cases} , \tag{6.6}$$

where $\Gamma = \sigma_\epsilon^2 \sum_{k=0}^{\infty} A^k 1_m (A^k)^\tau$, and where 1_m is a $p \times p$ matrix with 1 in the upper left corner and zeros elsewhere. Moreover, for $t > s$, $\mathrm{E}\,(y_{(i)t} y_{(j)s}^\tau) = A^{t-s} \mathrm{E}\,(y_{(i)t} y_{(j)t}^\tau)$.

By (6.3), (2.6), (6.6) and a reasoning analogous to that of Section 2 we have

$$\begin{aligned}
\operatorname{cov}(\tilde{a}) &\sim \mathrm{E}\left\{(\tilde{a}-a)(\tilde{a}-a)^{\tau}\right\} \\
&\sim \left(1-\frac{1}{n}\right)^{-1}\Gamma^{-1}\frac{1}{n^2(T-p)^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{t=p}^{T-1}\sum_{s=p}^{T-1}\mathrm{E}\left(u_{(i)t+1}u_{(j)s+1}y_{(i)t}y_{(j)s}^{\tau}\right)\left(1-\frac{1}{n}\right)^{-1}\Gamma^{-1} \\
&\sim \frac{\sigma_{\epsilon}^2\Gamma^{-1}}{(n-1)(T-p)} \qquad (6.7)
\end{aligned}$$

as $nT \to \infty$.

If T is small, again it is advantageous to use Burg-type estimates. In the univariate time series case the Burg-type estimates are built up recursively (cf. Robinson and Treitel 1980, appendix 16-2). The analogy in the panel case is given by $\tilde{a}_B^{\tau} = [\tilde{a}_{B,p,1}, \ldots, \tilde{a}_{B,p,p}]$ with

$$\tilde{a}_{B,p,p} = \frac{2\sum_{i=1}^{n}\sum_{t=1}^{T-p} b_{p(i)t}b'_{p(i)t}}{\sum_{i=1}^{n}\sum_{t=1}^{T-p}(b_{p(i)t}^2 + b_{p(i)t}'^2)} \qquad (6.8)$$

where

$$b_{p(i)t} = \begin{cases} Y_{(i)t}, & p = 1 \\ b_{p-1(i)t} - \tilde{a}_{B,p-1,p-1}b'_{p-1(i)t}, & p > 1 \end{cases} \qquad (6.9)$$

$$b'_{p(i)t} = \begin{cases} Y_{(i)t+1}, & p = 1 \\ b'_{p-1(i)t+1} - \tilde{a}_{B,p-1,p-1}b_{p-1(i)t+1}, & p > 1 \end{cases} \qquad (6.10)$$

and with

$$\tilde{a}_{B,p,k} = \tilde{a}_{B,p-1,k} + \tilde{a}_{B,p,p}\tilde{a}_{B,p-1,p-k}, \quad k = 1, \ldots, p-1, \quad p > 1. \qquad (6.11)$$

As $T \to \infty$, $\tilde{a}_B$ and $\tilde{a}$ have the same properties. For a small T, when $\tilde{a}_B$ is of most interest, it is difficult to compute the covariance matrix $\operatorname{cov}(\tilde{a}_B)$ in analogy to (3.12) directly from the above recursive formulae. The Burg-type estimate can be derived nonrecursively as the conditional least squares estimate resulting from minimizing a combined forecast and hindsight error, and from this representation it is at least in principle possible to find an asymptotic expression for $\operatorname{cov}(\tilde{a}_B)$. Here we will satisfy ourselves by presenting some simulations at the end of the section.

Next we introduce the estimate $\hat{a}^\tau = [\hat{a}_1, \ldots, \hat{a}_p]$ corresponding to $\hat{a}$ of (4.2). We define

$$\begin{aligned}\hat{a}^\tau &= \frac{1}{n(T-p)} \sum_{i=1}^{n} \sum_{t=p}^{T-1} (X_{(i)t+1} - \overline{X}_{\cdot\cdot,1})(x_{(i)t} - \overline{x}_{\cdot\cdot,0})^\tau \\ &\quad \times \left\{ \frac{1}{n(T-p)} \sum_{i=1}^{n} \sum_{t=p}^{T-1} (x_{(i)t} - \overline{x}_{\cdot\cdot,0})(x_{(i)t} - \overline{x}_{\cdot\cdot,0})^\tau \right\}^{-1} \qquad (6.12)\end{aligned}$$

with $\overline{X}_{\cdot\cdot,1} = n^{-1}(T-p)^{-1} \sum_{i=1}^{n} \sum_{t=p}^{T-1} X_{(i)t+1}$ and $\overline{x}_{\cdot\cdot,0} = n^{-1}(T-p)^{-1} \sum_{i=1}^{n} \sum_{t=p}^{T-1} x_{(i)t}$. Inserting from (6.1),

$$\begin{aligned}\hat{a}^\tau - a^\tau &= \frac{1}{n(T-p)} \sum_{i=1}^{n} \sum_{t=p}^{T-1} (\eta_{t+1} - \overline{\eta}_{\cdot,0} + \epsilon_{(i)t+1} - \epsilon_{\cdot,0})(x_{(i)t} - \overline{x}_{\cdot\cdot,0})^\tau \\ &\quad \times \left\{ \frac{1}{n(T-p)} \sum_{i=1}^{n} \sum_{t=p}^{T-1} (x_{(i)t} - \overline{x}_{\cdot\cdot,0})(x_{(i)t} - \overline{x}_{\cdot\cdot,0})^\tau \right\}^{-1}.\end{aligned}$$

Assuming that the $\{\eta_t\}$ are iid , we have as $T \to \infty$,

$$\Gamma_x = \mathrm{cov}\,(x_{(i)t}) \sim \mathrm{E}\,(x_{(i)t} - \overline{x}_{\cdot\cdot,0})(x_{(i)t} - \overline{x}_{\cdot\cdot,0})^\tau \sim (\sigma_\eta^2 + \sigma_\epsilon^2) \sum_{k=0}^{\infty} A^k 1_m (A^k)^\tau$$

so that $\Gamma_x = (\sigma_\eta^2 + \sigma_\epsilon^2)\Gamma/\sigma_\epsilon^2$. Moreover, using the same reasoning as in Section 4, for $i \neq j$,

$$\mathrm{E}\,\left\{(x_{(i)t} - \overline{x}_{\cdot\cdot,0})(x_{(j)t} - \overline{x}_{\cdot\cdot,0})^\tau\right\} \sim \sigma_\eta^2 \sum_{k=0}^{\infty} A^k 1_m (A^k)^\tau = \rho\Gamma_x$$

and hence as $T \to \infty$,

$$\begin{aligned}\mathrm{cov}\,(\hat{a}) &\sim \frac{1}{n(T-p)} \Gamma_x^{-1} \left\{(\sigma_\eta^2 + \sigma_\epsilon^2)\Gamma_x + (n-1)\sigma_\eta^2 \rho \Gamma_x\right\} \Gamma_x^{-1} \\ &= \frac{(\sigma_\eta^2 + \sigma_\epsilon^2)\Gamma_x^{-1}\left\{1 + (n-1)\rho^2\right\}}{n(T-p)} = \frac{\sigma_\epsilon^2 \Gamma^{-1}\left\{1 + (n-1)\rho^2\right\}}{n(T-p)}. \qquad (6.13)\end{aligned}$$

Comparing (6.7) and (6.13), we get the same cross-over point at $\rho = 1/(n-1)$ as for $p = 1$ in Section 4. This holds both for Burg and non-Burg estimates as the asymptotics are the same as $T \to \infty$. For a finite T it is checked for the non-Burg estimates by simulation experiments in Table 7, which is based on 5000 realizations of the model

$$X_{(i)t} = X_{(i)t-1} - 0.6X_{(i)t-2} + 0.2X_{(i)t-3} - 0.2X_{(i)t-4} + 0.4X_{(i)t-6} + \eta_t + \epsilon_{(i)t} \qquad (6.14)$$

with $T = 1000$ and $T = 100$.

As in the first order autoregressive case, the Burg-type estimators $\hat{a}_B$ and $\tilde{a}_B$ have less mean square error when T is small. Figure 6 shows the empirical distribution function for $\hat{a}_{B,3,1}$, $\hat{a}_{B,3,2}$ and $\hat{a}_{B,3,3}$ defined in analogy with (6.8) - (6.11), with $Y_{(i)t}$ and $Y_{(i)t+1}$ replaced by $X_{(i)t} - \overline{X}_{\cdot\cdot,0}$ and $X_{(i)t+1} - \overline{X}_{\cdot\cdot,0}$, respectively, for different values of ρ and the corresponding non-Burg estimators $\hat{a}_1, \hat{a}_2$ and $\hat{a}_3$ given by (6.12) (lines with bullets) in the uncorrelated case. The model used is

$$X_{(i)t} = -0.5X_{(i)t-1} + 0.5X_{(i)t-3} + \eta_t + \epsilon_{(i)t} \tag{6.15}$$

with $T = 4$ and $n = 128$. The poor behavior for $\rho > 0$ is of course due to the inconsistency of the $\hat{a}$-estimates in this situation.

7 A real data example

We end by taking a closer look at the grey-sided vole data depicted in Figure 1, which contains $n = 41$ series, each consisting of $T = 31$ observations. There are indications of a weak nonlinearity in the data, so that our results, using linear models, are approximate. Looking at the plots of Figure 1, there are clear signs of intercorrelation. We have estimated the intercorrelation ρ by

$$\hat{\rho} = \frac{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sum_{t=1}^{T}(X_{(i)t} - \overline{X}_{(i)\cdot})(X_{(j)t} - \overline{X}_{(j)\cdot})}{\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left\{\sum_{t=1}^{T}(X_{(i)t} - \overline{X}_{(i)\cdot})^2 \sum_{t=1}^{T}(X_{(j)t} - \overline{X}_{(j)\cdot})^2\right\}^{1/2}} = 0.437 \tag{7.1}$$

where $\overline{X}_{(i)\cdot} = T^{-1}\sum_t X_{(i)t}$. Another alternative is to take the average of the ordinary correlation between all possible pairs of series. This average correlation was computed to 0.433. In any case the estimated value of ρ is very much larger than the threshold value of $\rho = 1/(n-1) = 1/40 = 0.025$, which means that we put considerable more trust in the estimates $\tilde{a}$ and $\tilde{a}_B$ developed in Section 3 than in the estimates $\hat{a}$ and $\hat{a}_B$ described in Section 4.

We have assumed that the order of the autoregressive model is known in this paper. For a real data set the order must be determined, but we will postpone the systematic investigation of an order determination procedure to a later publication. If we assume that the observations follow a first, second or third order model, respectively, the corresponding coeffisient estimates are given in Table 8. In the first order case the 95 % estimated confidence intervals are given as well. These are obtained by using (2.11) and (4.10) with a and ρ replaced by estimated values. For the variance of $\tilde{a}$ and $\hat{a}$ we get 0.00082 and 0.00701, respectively. The confidence interval for $\tilde{a}$ is more reliable than that of $\hat{a}$, because the latter is based on (4.10) which requires T to be large and $\{\eta_t\}$ to consist of iid variables. Since T is not very small, we cannot expect Burg type estimators $\tilde{a}_B$ and $\hat{a}_B$ to be very different from $\tilde{a}$ and $\hat{a}$, and this is comfirmed by Table 8. But whether we use $\hat{a}$ or $\tilde{a}$ very definitely makes a difference, which suggests that care should be exercised in the choice of autoregressive coefficient estimates in a panel situation.

We have estimated the common effect process η_t by

$$\hat{\eta}_t = \overline{X}_{\cdot t} - \tilde{a}_1 \overline{X}_{\cdot t-1} - \tilde{a}_2 \overline{X}_{\cdot t-2} - \tilde{a}_3 \overline{X}_{\cdot t-3}.$$

There are only $31 - 3 = 28$ observations available to judge the properties of $\hat{\eta}_t$, but computing the autocorrelation of $\hat{\eta}_t$ does suggest that this process may be autocorrelated, which gives one more reason for prefering $\tilde{a}$ to $\hat{a}$, as the former is not sensitive to the properties of $\{\eta_t\}$, whereas the latter requires $\{\eta_t\}$ to be iid in order for it to be consistent.

Acknowledgement

We are grateful to T. Saitoh and N.C. Stenseth for making the grey sided vole data available to us, and we are indebted to E. Biørn, M. Hallin, L. Reichlin, H. Tong and Q. Yao for useful comments.

References

BALTAGI, B. H. (1995). *Econometric Analysis of Panel Data.* New York: John Wiley & Sons.

Bjørnstad, O. N., Champley, S., Stenseth, N. C. and Saitoh, T. (1996). Cyclicity and stability of grey-sided voles, Clethrionomys rufocanus, of Hokkaido: spectral and principal components analyses. *Phil. Trans. R. Soc. Lond.* B **351**, 867-75.

Bloomfield, P., Oehlert, G., Thompson, M. L. and Zeger, Z. (1983) A frequency domain analysis of trends in Dobson total ozone records. *J. Geophys. Res.*, **88**, 8512-22.

Brillinger, D. R. (1973). The analysis of time series collected in experimental design. In *Multivariate analysis III* (ed. P. R. Krishnaiah), pp. 241-56. New York: Academic Press.

Brillinger, D. R. (1980) Analysis of variance and problems under time series analysis. In *Handbook of Statistics* (ed. P. R. Krishnaiah), vol. 1, pp. 237-78. Amsterdam: North Holland.

Brockwell, P. J. and Davis, R. A. (1996). *Introduction to Time Series and Forecasting.* New York: Springer-Verlag.

Cox, D. R. and Solomon, P. J. (1988) On testing for serial correlation in large numbers of small samples. *Biometrika*, **75**, 145-8.

Diggle, P. J., Liang, K. and Zeger, S. L. (1994). *Analysis of Longitudinal Data.* Oxford: Clarendon Press.

Diggle, P. J. and Al Wasel, I. (1997). Spectral analysis of replicated biomedical time series (with discussion). *Appl. Statist.*, **46**, 31-71.

Forni, M. and Reichlin, L. (1997). Let's get real: a factor analytical approach to disaggregated business cycle demands. Preprint, University of Bruxelles (ECARE).

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and its Application.* New York: Academic Press.

HOLTZ-EAKIN, D., NEWEY, W. AND ROSEN, H. (1988). Estimating vector autoregressions with panel data. *Econometrica*, **75**, 1371-95.

HSIAO, C. (1986). *Analysis of Panel Data.* Cambridge: Cambridge University Press.

MÁTYÁS, L. AND SEVESTRE, P. (1992). *The Econometrics of Panel Data.* Dordrecht: Kluwer Academic Publishers.

ROBINSON, E. A. AND TREITEL, S. (1980). *Geophysical Signal Analysis.* Englewood Cliffs: Prentice-Hall.

STENSETH, N. C., BJØRNSTAD, O. N. AND SAITOH, T. (1996). A gradient from stable to cyclic populations of Clethrionomys rufocanus in Hokkaido, Japan. *Proc. R. Soc. Lond.* B **263**, 1117-26.

Figure captions

Figure 1. The figure shows $\log(X_{(i)t} + 1)$ where $\{X_{(i)t}, i = 1, \ldots, 41, t = 1, \ldots, 31\}$ is the number of grey-sided voles trapped each year from 1962 to 1992 in 41 different locations in Hokkaido, Japan.

Figure 2. The figure is based on 5000 realisations of model (2.1) with $n = 128$, $T = 2$, $\eta_t \equiv 0$ and $\sigma_\epsilon^2 = 1$. It shows the empirical distribution function of $\tilde{a}_B$ and $\tilde{a}$ for different values of the coefficient a.

Figure 3. The figure is based on 5000 realisations of model (2.1) with $a = 0.5$ and different combinations of n and T. It shows the empirical distribution function of $\tilde{a}_B$ and of $\hat{a}_B$ for $\rho = 0.0$, 0.5 and 0.9. The simulated means are drawn in as vertical bars.

Figure 4. The figure is based on 5000 realisations of model (2.1) with $n = 128$, $a = 0.5$ and $\rho = 0.5$. It shows for different values of T the empirical distribution function of $\hat{a}_B$ and $\tilde{a}_B$. The simulated means are drawn in as vertical bars.

Figure 5. The figure is based on $N = 5000$ realisations of model (2.1) with $T = 2$, $a = 0.5$ and $\rho = 0.5$. It shows for different values of n a density estimate of $\tilde{a}$ and of $\tilde{a}_B$ (thick lines), compared to the $\mathcal{N}(\hat{\mu}, \hat{\sigma}^2)$-density, where $\hat{\mu}$ and $\hat{\sigma}^2$ is the simulated mean and variance of $\tilde{a}$ and $\tilde{a}_B$, respectively. A kernel estimator with bandwidth $h = \hat{\sigma} N^{-1/5}$ is used.

Figure 6. The figure is based on 5000 realisations of model (6.15) with $n = 128$ and $T = 4$. It shows for $i = 1, 2, 3$ the empirical distribution function of $\hat{a}_{B,3,i}$ for $\rho = 0$ (thick lines), $\rho = 0.2$ and $\rho = 0.5$, and of $\hat{a}_i$ (lines with bullets) for $\rho = 0$.

Table captions

Table 1. The table is based on 5000 realisations of model (2.1) with $\eta_t \equiv 0$ and $\sigma_\epsilon^2 = 1$. It shows for different values of a, n and T the simulated value $n(T-1)\widehat{\mathrm{var}}\,(\tilde{a}_B)$ and, in parentheses, the ratio $\widehat{\mathrm{var}}\,(\tilde{a}_B)/\mathrm{var}\,(\tilde{a}_B)$ where $\mathrm{var}\,(\tilde{a}_B)$ is given by (3.12).

Table 2. The table is based on the same simulations as Figure 2. It shows the simulated mean square errors of $\tilde{a}_B$ and of $\tilde{a}$, the ratio between them, and the asymptotic ratio between the variances of $\tilde{a}_B$ and $\tilde{a}$ given by (3.13).

Table 3. The table is based on the same simulations as Figure 3. It shows the simulated mean square errors of $\hat{a}_B$ and $\tilde{a}_B$. For $nT = 20000$ the last decimals of the asymptotic variance of $\hat{a}_B$ given by (4.10), and of $\tilde{a}_B$ given by (2.11) are given in parentheses.

Table 4. The table is based on 5000 realisations of model (2.1) with $a = 0.5$ and $nT = 20000$. It shows the simulated variance of $\hat{a}_B$ and $\tilde{a}_B$ compared to the asymptotic variance of $\tilde{a}_B$ for different values of ρ and n. For the three first columns, $\rho = 1/(n-1)$.

Table 5. The table is based on 5000 realisations of model (2.1) with $a = 0.5$.

Table 6. The table is based on the same simulations as Figure 4.

Table 7. The table is based on 5000 realisations of model (6.14). For the three first rows of each part of the table, $\rho = 1/(n-1)$.

Table 8. The table shows different estimates of the coefficients in an AR(p) approximation ($p = 1, 2, 3$) to the log-transformed grey-sided vole data of Figure 1. For $p = 1$ the estimated 95 % confidence limits are shown.

Figure 1:

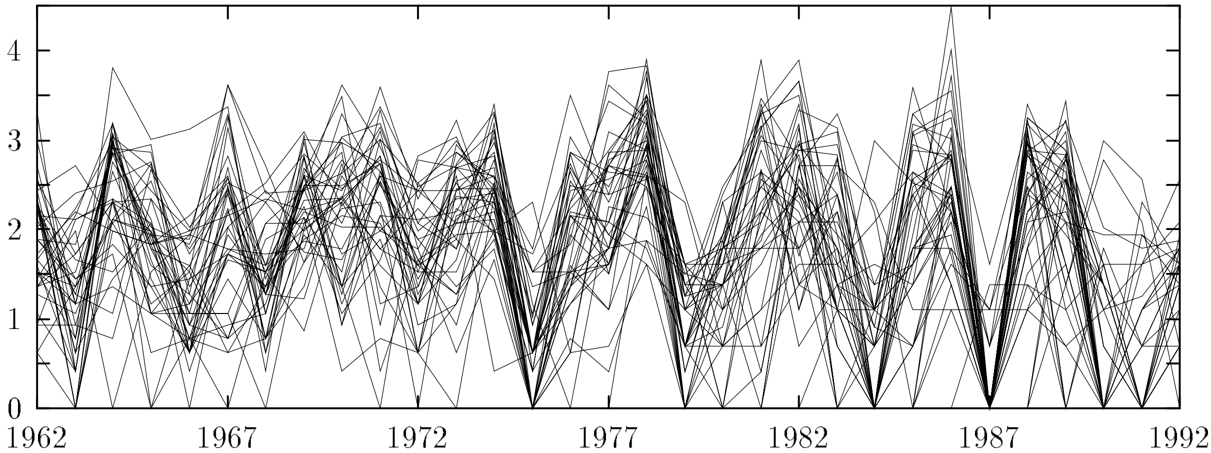

Figure 2:

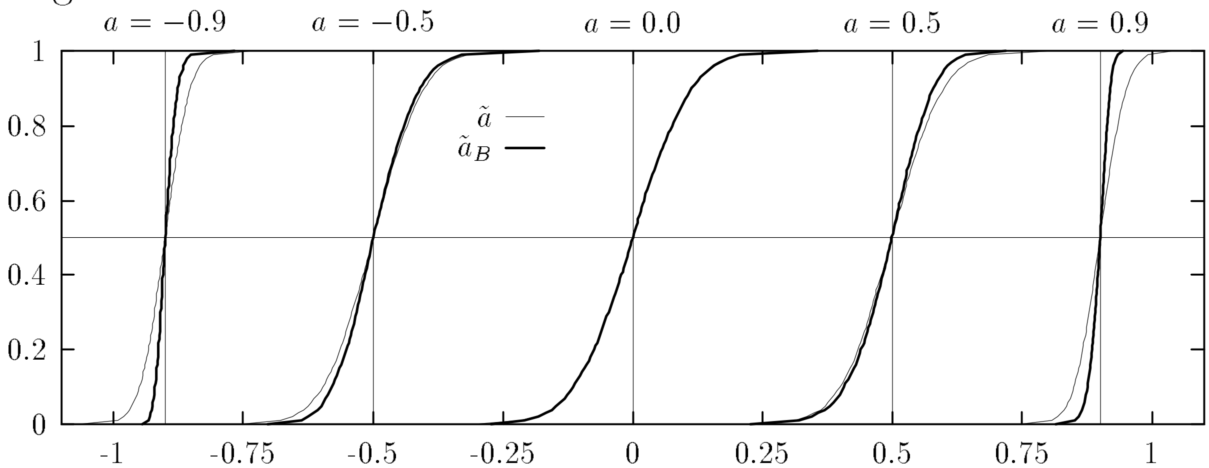

Figure 3: